Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 1979–1983, March 1982
Immunology

Single amino acid substitution altering antigen-binding specificity

(immunoglobulin/mutation/phosphocholine/antibody diversity)

STUART RUDIKOFF*, ANGELA M. GIUSTI†, WENDY D. COOK†, AND MATTHEW D. SCHARFF†

*Laboratory of Cell Biology, National Institutes of Health, Bethesda, Maryland; and †Department of Cell Biology, Albert Einstein College of Medicine, Bronx, New York

Communicated by Michael Potter, November 23, 1981

ABSTRACT **S107, a phosphocholine-binding myeloma protein, has been cloned in soft agar, and an antigen-binding variant has been isolated and characterized. The variant does not bind phosphocholine attached to carrier or as free hapten in solution but does retain antigenic determinants (idiotypes) of the parent. Chain recombination experiments suggest that the defect in binding is entirely in the heavy chain. Amino acid sequence analysis showed a single substitution—glutamic acid to alanine at position 35—in the first hypervariable or complementarity-determining region. In terms of the three-dimensional model of the phosphocholine-binding site, glutamic acid-35 provides a hydrogen bond to tyrosine-94 of the light chain that appears to be critical for stability of this portion of the binding site. The removal of this bond and the presence of the smaller alanine side chain is thus consistent with the loss in binding activity. These results suggest that small numbers of substitutions in antibodies, such as those presumably introduced by somatic mutation, may in *some* situations be effective in altering antigen-binding specificity.**

The generation of antibody diversity has long been and remains one of the intriguing questions in immunology. Protein sequence analyses (1, 2) and nucleic acid studies (3–7) have similarly suggested that the number of light (L) and heavy (H) chain genes in the germ line is large (>200 each). If random combinations of L and H chains were to occur, >10,000 different antibodies could be generated solely from the germ-line repertoire. Furthermore, immunoglobulin chains are encoded in multiple genetic elements. The variable domains of light chains are encoded by two gene segments designated variable (V) and joining (J) (8, 9). Heavy chain variable domains, in addition to V and J segments, have a third element, D (diversity), that encodes a portion of the third hypervariable region (10–12). The combination of a given L chain V gene with any of four functional J genes can thus produce additional structural diversity as can V, D, and J recombination in the H chain. The potential sequence diversity is further increased by variations in the sites at which these elements combine (8, 9, 13–17). At present, it is not clear how much the sequence diversity generated by these events contributes to functional changes that affect the specificity and affinity of antigen binding.

In view of the large amount of structural diversity that can be generated from the germ-line repertoire and the recombination events occurring during the formation of active immunoglobulin genes, the question of the occurrence and role of somatic mutation in the generation of antibody diversity remains to be determined. The initial studies of mouse λ L chains by Weigert and co-workers (18, 19) identified 12 invariant sequences, 5 with single amino acid substitutions, 1 with two substitutions, and 1 with three substitutions. All interchanges were located in hypervariable regions and it was concluded that the variants arose by a somatic mutation process. This interpretation has been confirmed by the finding that a single germ-line V region gene and J segment encode the λ_1 V regions (20). While the sequence changes in λ_1 L chains were confined to hypervariable regions, studies of a κ chain subgroup (V_{k21}) also suggested somatic mutations, but these substitutions were observed in framework as well as hypervariable residues (21). Most relevant to the studies described here is a recent report by Gearhart *et al.* (22) on the structure of H and L chains from phosphocholine (*P*-Cho)-binding hybridomas. A number of sequences were determined for IgM and IgG hybridomas and were compared with previously determined structures from IgA myeloma proteins. All of the μ chains were found to be identical in the V region, while all of the γ chains and about half of the α chains had substitutions that occurred in both hypervariable and framework regions. An analysis of the germ-line genes encoding the *P*-Cho heavy chains (23) suggested that all except one of the *P*-Cho heavy chains were encoded by a single gene that faithfully encodes the T15 IgA myeloma V region, as well as all of the μ V regions. These and similar findings among H chains from antinitrophenyl antibodies (24) argue that somatic mutation is occurring and that it is not restricted to hypervariable regions.

To date, it is not known whether the limited number of amino acid substitutions presumably generated by somatic mutation can be effective in altering antigen binding specificity or affinity. We have attempted to approach this question by examining the structure of antigen-binding variants derived from the *P*-Cho-binding myeloma protein S107. S107 is a BALB/c IgA, κ myeloma protein that appears to be identical in structure to the T15 myeloma protein and ≈90% of the BALB/c antibodies elicited by immunization with *P*-Cho-containing antigens (25). We have chosen this sytem because of the availability of a number of primary sequences from *P*-Cho-binding myeloma proteins and a three-dimensional structure of the Fab fragment from the closely related protein M603 (26, 27). Antigen-binding variants were selected by cloning S107 cells in soft agar and overlaying with *P*-Cho–keyhole limpet hemocyanin (KLH) (28). Clones that are not surrounded by antigen–antibody precipitates and occur at a spontaneous frequency of 0.1–1% are presumed to be antigen-binding variants. One of these first generation variants, U4, has been characterized in detail and is the subject of the present communication.

MATERIALS AND METHODS

The S107 cell line was originally obtained from the Salk cell bank. Propagation of the cells, their cloning, and identification

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: L and H, light and heavy chain genes or their respective polypeptide chains; V, J, and D, variable, joining, and diversity (region), respectively; *P*-Cho, phosphocholine; KLH, keyhole limpet hemocyanin; SRBC, sheep erythrocytes.

of variants have been described previously (28). *P*-Cho was attached to sheep erythrocytes (SRBC) (29), and hemagglutination of *P*-Cho–SRBC was determined as described by Evans *et al.* (30). Radioimmunoassays were carried out by the method of Pierce and Klinman (31) in which antigen (*P*-Cho–KLH) or antibody (28) was attached to polyvinyl plates (Dynatech). The S107 immunoglobulin and the rat anti-S107 monoclonal antibody were biosynthetically labeled by incubating the cultured myeloma cells with [^{35}S]methionine. Labeling of the antibodies and their use in the radioimmunoassay have been described (32). Equilibrium dialysis was carried out by the flow dialysis method of Colowick and Womack (33).

S107 and variant immunoglobulins were purified from the ascites of tumor-bearing animals. S107 was purified by affinity chromatography on *P*-Cho-Sepharose. The U4 protein was chromatographed on DEAE-Sephacel (Pharmacia) and then further purified by gel filtration on Sephacryl S200 (Pharmacia). Chain recombination was carried out as described by Manjula *et al.* (34) except that the denaturing buffer was 5 M guanidine in 0.2 M NH_4HCO_3.

For sequence studies, the H chains of both the S107 and U4 immunoglobulin were isolated on Sephadex G-100 columns equilibrated in 6 M urea/1 M HOAc. H chains were cleaved with CNBr and the fragments were isolated by gel filtration (35). Sequences were determined on a modified (36, 37) Beckman 890C sequencer using a Quadrol buffer system as described (38).

RESULTS AND DISCUSSION

Characterization of the U4 Variant. When a freshly isolated subclone of the S107 cell line is recloned in soft agar and overlayed with *P*-Cho–KLH, 0.1–1% of the clones are not surrounded by a visible antigen–antibody precipitate (28). Between 50% and 70% of these presumptive variant clones continue to secrete IgA molecules in the same amount as the parental clone. Most of these variant antibodies have normal-sized H and L chains and are polymeric molecules that are indistinguishable from the parental IgA (28, 39). One of these variants, U4, was grown to mass culture and injected into BALB/c mice, and the variant protein was purified from ascites fluid by DEAE-Sephacel and Sephacryl chromatography. Comparison of the antigen and hapten binding of the U4 variant and the parental S107 protein is presented in Table 1. U4 does not agglutinate *P*-Cho–SRBC and does not bind hapten when assayed by equilibrium dialysis. To further evaluate the antigen binding capacity of U4, *P*-Cho–KLH was absorbed to the walls of polyvinyl microtiter plates and unlabeled U4 and S107 protein were compared for their ability to compete with endogenously labeled S107 for antigen. U4 demonstrated minimal binding, only slightly more than an unrelated antibody, in competing for *P*-Cho in this assay. These results indicated that the U4 variant had lost its ability to bind antigen and hapten. When parent and U4 were analyzed for antigenic determinants (idiotypes) detected by a particular monoclonal antibody (Fig. 1),

Table 1. Antigen- and hapten-binding characteristics of myeloma S107 and variant U4

Cell line	Hemagglutination titer,* *P*-Cho–SRBC	Radioimmunoassay,† %*P*-Cho–KLH	K_a, 10^5 M^{-1}
S107.3.4	8192	100	2.3
U4	0	<0.01	0

* Purified protein (250 ng) was examined in 1:2 dilutions.
† % relative binding = (ng of parent required for 50% inhibition/ng of variant required for 50% inhibition) ×100.

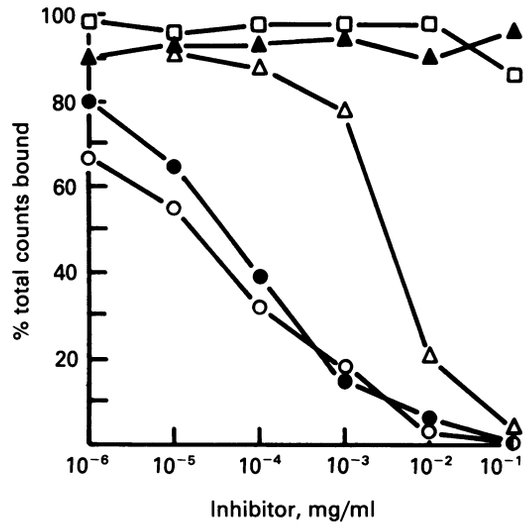

FIG. 1. Competition radioimmunoassay of S107 (●) and U4 (○) with a rat monoclonal antibody raised against S107. Purified S107 protein was attached to polyvinyl plates and parent and variant were incubated with 50% saturating amounts of [^{35}S]methionine-labeled rat monoclonal antibody in the presence of various amounts of the proteins indicated. M603 (▲) and M511 (△) are *P*-Cho-binding myeloma proteins. Reagents recognizing S107 frequently crossreact with M511. W3129 (□) does not bind *P*-Cho.

U4 was found to react with this reagent to approximately the same extent as the parent. By using conventional antisera to the *P*-Cho-binding site and a variety of additional antivariable region monoclonal antibodies, we have observed that some distinguish these two proteins to the extent that U4 reacts ≈1/10th as well as the parent while others fail to discriminate between parent and variant. Thus, the association of idiotype with a non-antigen-binding molecule (U4) is likely to be highly dependent on the particular reagent used. Taken together, the above data indicate that an alteration has occurred in the configuration of the U4 binding site that has drastically altered antigen-binding specificity but only minimally altered antigenic determinants.

To more precisely define the defect in U4, chain recombination experiments were performed between parent and variant. H and L chains were isolated from partially reduced proteins, mixed together in various combinations, and reassociated by removing the denaturing reagent (Table 2). The recombinants containing the H chain of U4 and the L chain of S107 did not hemagglutinate *P*-Cho–SRBC. However, the combination of the parental H chain and the U4 L chain not only agglutinated *P*-Cho–SRBC (Table 2) but also had the same affinity for *P*-Cho by equilibrium dialysis as the S107 protein (data not shown). The loss in antigen-binding activity was therefore associated entirely with the U4 heavy chain.

Structural Difference Between Parent and Variant. Since alterations in antigen binding were found to be associated with the U4 H chain, we have determined the primary structure of the entire H chain variable region of the variant and parental proteins. A single amino acid substitution—alanine for glutamic acid—was found at position 35 in the first hypervariable or complementarity-determining region (Fig. 2). This interchange can be explained by a single-base substitution in the second nucleotide position of the glutamic acid codon or could be due to a two-base change. A significant association between *P*-Cho binding and the occurrence of phenylalanine-32, tyrosine-33, and glutamic acid-35 has previously been noted (41). The effect of the observed alanine–glutamic acid interchange is striking in

Table 2. Chain recombination of S107 and U4

	Hemagglutination titer
$H^{107} + L^{107}$	10
$H^{107} + L^{U4}$	7.5
$H^{U4} + L^{107}$	1
$H^{U4} + L^{U4}$	1

Chain recombination was as described in *Materials and Methods*. The recombinants were not further purified. Equal amounts of protein, which also contained the same amount of idiotype by inhibition of hemagglutination, were allowed to react with *P*-Cho–SRBC. Facilitating rabbit anti-IgA antibody was added and the hemagglutinin titer is expressed as the square of the logarithm.

terms of the model of the *P*-Cho-binding site of M603 (Fig. 3). M603 is a *P*-Cho-binding myeloma protein whose three-dimensional structure has been determined (26). Crystallographic analysis has indicated that the binding of *P*-Cho hapten is almost entirely associated with amino acids in the H chain. When the H chains of M603 and S107 are compared, five differences are observed (four substitutions plus a one amino acid size difference), none of which occur at positions identified as participating in hapten contact. Although the L chains from M603 and S107 are from different V_κ subgroups and hence have considerable sequence variation (42–44), the one amino acid (leucine-96) directly implicated in hapten contact (26, 43) is also preserved in the S107 L chain. Furthermore, anti-*P*-Cho hybridomas have been identified (22) that are comprised of an M603 L chain paired with a T15 H chain and have affinities for *P*-Cho similar to that of the M603 myeloma protein (45).

Based on the above considerations as well as predictive calculations, we have concluded that the principles derived from the structure of the M603-binding site will largely be applicable to the S107 site (46). In the M603 structure, glutamic acid-35 was originally designated as a hapten-contacting residue (26). From continued refinement of the x-ray structure analysis, it now appears that the distance between *P*-Cho and glutamic acid-35 makes its role in direct hapten contact somewhat uncertain (D. R. Davies, personal communication). However, from Fig. 3, it can be seen that the side chain of glutamic acid-35 is hydrogen bonded to the phenolic hydroxyl of the L chain tyrosine-94 (43). This bond appears to be important in stabilizing this portion of the binding pocket that contains the hapten-contacting residues tyrosine-33 and arginine-52 from the H chain and leucine-96 from the L chain. In addition, the "vacuum" created by the removal of the carboxyl and methyl groups of glutamic acid would further distort this region. In fact, if glutamic acid-35 were only a hapten-contacting residue, it might be expected that substitution of the smaller alanine side chain would merely produce a decrease in affinity. This substitution, while possibly having only a minimal effect on hapten contact, apparently removes a bond and side chain essential to the structural integrity of this portion of the binding site. It is interesting to note that all H chains from conventionally induced anti-*P*-Cho antibodies are similar in amino acid sequence. From analysis of the gene family encoding the *P*-Cho H chain V regions (23), it can be seen that only one of three potentially expressed genes encodes glutamic acid at position 35. It is this gene that directly translates into the S107 protein sequence and presumably encodes most other *P*-Cho H chains. Thus, the predominant use of this particular gene in anti-*P*-Cho antibodies appears to correlate with the presence of glutamic acid-35. The two other genes code for serine at position 35 (as well as several additional differences), but the remaining major hapten-contacting residues are also present in these genes. One exception has been observed to date in that a single hybridoma (22) has been described that has a serine at position 35 and other substitutions characteristic of the two "unused" genes. This protein has the lowest affinity for *P*-Cho of any hybridoma or myeloma characterized thus far. It is thus possible that the presence of serine at residue 35 may maintain enough of the needed conformation to permit *P*-Cho binding (albeit at a lower affinity) in contrast to the complete loss associated with the alanine substitution in

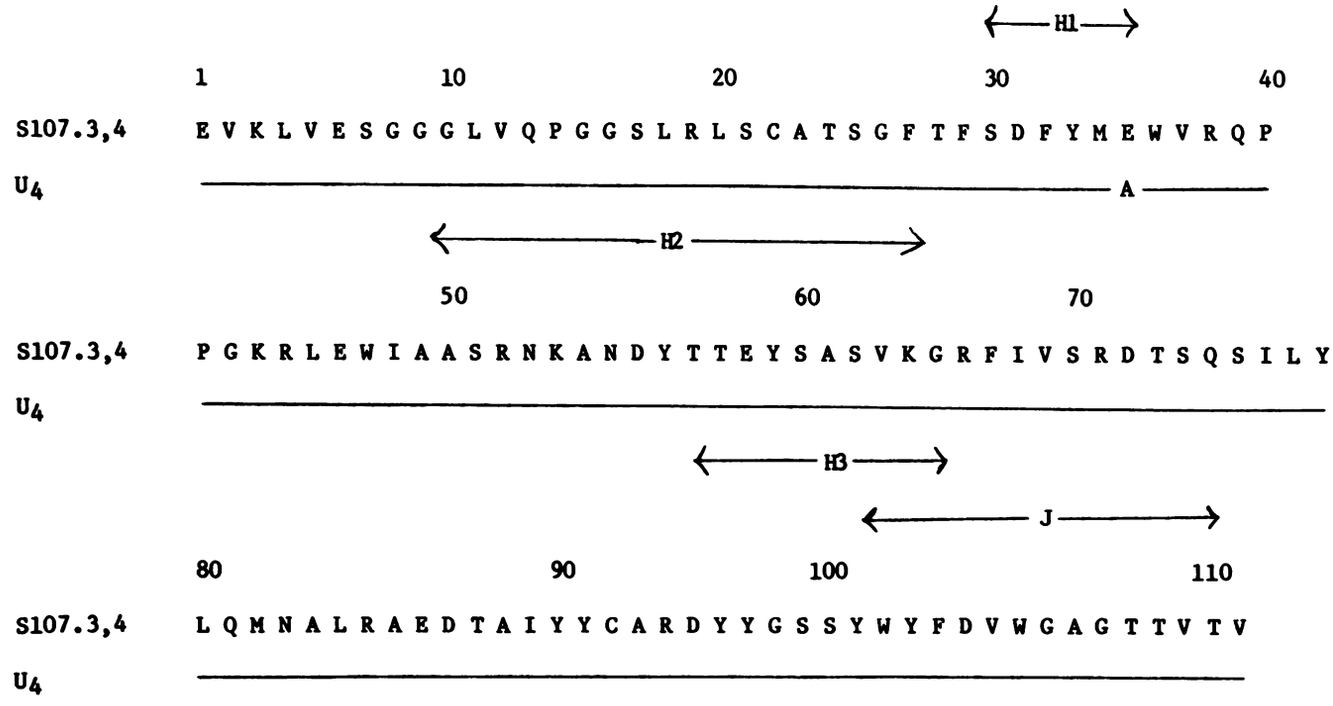

FIG. 2. Amino acid sequence of the H chain variable region of parent and variant. Cyanogen bromide fragments were isolated and their sequences were determined as described (35). Numbering is according to Kabat *et al.* (40).

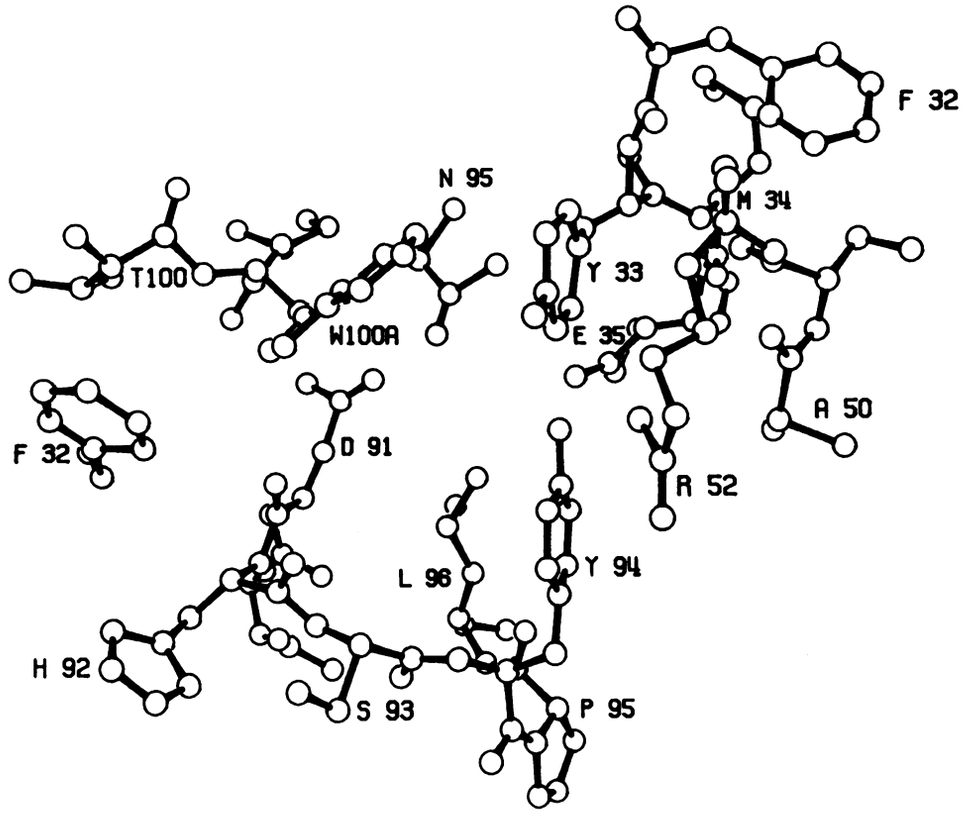

FIG. 3. Schematic representation of the *P*-Cho-binding site of M603 taken from ref. 43. H chain complementarity-determining regions are located in the upper portion. L chain residues contributing to the binding site occur mainly in the third L chain complementarity-determining region (lower portion). The effects of substitution in U4 are loss of the hydrogen bond between glutamic acid (E)-35 and tyrosine (Y)-94 and the accompanying decrease in side chain volume that appear to be critical to stability of this part of the site.

U4. Alternatively, the additional substitutions in this chain may partially compensate for the replacement of glutamic acid-35.

Although the H chains from *P*-Cho-binding antibodies are similar in amino acid sequence, L chains from three different V region groups that vary greatly in sequence may be associated with these H chains (42). However, all three of these L chains use the J5 joining segment and have the L chain sequence tyrosine-94, proline-95, leucine-96 (38, 43, 44), so that the bond between glutamic acid 35H and tyrosine 94L can be formed with any of these L chains. The sequence Tyr-Pro-Leu at positions 94, 95, and 96 is found only in L chains from *P*-Cho-binding proteins.

Implications for Generation of Diversity. We have shown that a single amino acid substitution is capable of completely altering antigen-binding specificity. Thus, a small number of amino acid substitutions, such as those postulated to arise by somatic mutation, can potentially be effective in generating antibody diversity in addition to that inherent in the germ-line repertoire. The high spontaneous frequency of mutants (28) observed for the generation of non-antigen-binding variants in our system further suggests that events of this nature are not infrequent. We have characterized another primary variant of S107 that has decreased antigen binding and a single amino acid substitution in the fifth residue of its J segment (39). However, it is clear that all such substitutions need not and probably do not affect antigen binding. For example, the heavy chain from the *P*-Cho-binding myeloma protein M167 (35) differs from that of S107 at 13 positions (8 in hypervariable regions including a size difference) and yet has an association constant for hapten only slightly lower than S107. We have previously shown that, among anti-1,6-galactan-binding myeloma proteins, as many as eight or nine substitutions may occur in hypervariable regions with no significant effect on hapten affinity or specificity (13). Since these systems, as is the case of most hybridoma systems being examined today, are positively selected by antigen, they will in general reveal only substitutions not producing large changes in antigen binding. The negative selection used in this study permits analysis of changes that produce important phenotypic binding variation. Characterization of additional variants in this system may further define the potential of somatic mutations to introduce changes in antigen-binding specificity.

We thank Dr. M. W. Hunkapiller for his invaluable help in sequencer modifications. We also thank Terry Kelly and Mary Heller for their expert technical assistance. This work was supported by grants from the National Institutes of Health (AI 10702 and AI 5231), the National Science Foundation (PCM 77-25635), and the American Cancer Society (NP-317). A.M.G. is a Predoctoral Trainee supported by the National Cancer Institute (5T32-CA09172-06).

1. Hood, L., Loh, E., Hubert, J., Barstad, P., Eaton, B., Early, P. W., Fuhrman, J., Johnson, N., Kronenberg, M. & Schilling, J. (1976) *Cold Spring Harbor Symp. Quant. Biol.* **41**, 817–836.
2. Potter, M. (1977) *Adv. Immunol.* **25**, 141–211.
3. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. (1978) *Science* **202**, 11–17.
4. Lenhard-Schuller, R., Hohn, B., Brack, C., Hirama, M. & Tonegawa, S. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4709–4713.
5. Cory, S. & Adams, J. M. (1980) *Cell* **19**, 37–51.
6. Rabbitts, T. H., Matthyssens, G. & Hamlyn, P. H. (1980) *Nature (London)* **284**, 238–243.
7. Davis, M. M., Calame, K., Early, P. W., Livant, D. L., Joho, R., Weissman, I. L. & Hood, L. (1980) *Nature (London)* **283**, 733–739.

8. Max, E., Seidman, J. G. & Leder, P. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 3450–3454.
9. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. (1979) *Nature (London)* **280,** 288–294.
10. Early, P. W., Huang, H., Davis, M. M., Calame, K. & Hood, L. (1980) *Cell* **19,** 981–992.
11. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. (1980) *Nature (London)* **286,** 676–683.
12. Sakano, H., Kurosawa, Y., Weigert, M. & Tonegawa, S. (1981) *Nature (London)* **290,** 562–565.
13. Rudikoff, S., Rao, D. N., Glaudemans, C. P. J. & Potter, M. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 4270–4274.
14. Max, E., Seidman, J. G., Miller, H. & Leder, P. (1980) *Cell* **21,** 793–799.
15. Weigert, M., Perry, R., Kelley, D., Hunapiller, T., Schilling, J. & Hood, L. (1980) *Nature (London)* **283,** 497–499.
16. Gough, N. M. & Bernard, O. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 509–513.
17. Kwan, S. P., Max, E., Seidman, J. G., Leder, P. & Scharff, M. D. *Cell* **26,** 57–66.
18. Weigert, M., Cesari, I. M., Yonkovich, S. J. & Cohn, M. (1970) *Nature (London)* **228,** 1045–1047.
19. Cesari, I. M., & Weigert, M. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 2112–2116.
20. Bernard, O., Hozumi, N. & Tonegawa, S. (1978) *Cell* **15,** 1133–1144.
21. Weigert, M., Gatmaitan, L., Loh, E., Schilling, J. & Hood, L. (1978). *Nature (London)* **291,** 29–34.
22. Gearhart, P., Johnson, N. D., Douglas, R. & Hood, L. (1981) *Nature (London)* **291,** 29–34.
23. Crews, S., Griffin, J., Huang, H., Calame, K. & Hood, L. (1981) *Cell* **25,** 59–66.
24. Bothwell, A., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. & Baltimore, D. (1981) *Cell* **24,** 625–637.
25. Kohler, H. (1975) *Transplant. Rev.* **27,** 24–56.
26. Segal, D. M., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M. & Davies, D. R. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 4298–4302.
27. Padlan, E. A., Davies, D. R., Rudikoff, S. & Potter, M. (1976) *Immunochemistry* **13,** 945–949.
28. Cook, W. D. & Scharff, M. D. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5687–5691.
29. Chesebro, B. & Metzger, H. (1972) *Biochemistry II* **11,** 766–771.
30. Evans, J., Steel, M. & Arthur, E. (1974) *Cell* **3,** 153–158.
31. Pierce, S. & Klinman, N. (1976) *J. Exp. Med.* **144,** 1254–1262.
32. Yelton, D. E., Desaymard, C. & Scharff, M. D. (1981) *Hybridoma* **1,** 5–11.
33. Colowick, S. P. & Womack, F. C. (1969) *J. Biol. Chem.* **244,** 774–777.
34. Manjula, B. N., Glaudemans, C. P. J., Mushinski, E. B. & Potter, M. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 932–936.
35. Rudikoff, S. & Potter, M. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 2109–2112.
36. Wittmann-Liebold, B., Graffunder, H. & Kohls, H. (1976) *Anal. Biochem.* **75,** 621–633.
37. Hunkapiller, M. W. & Hood, L. (1978) *Biochemistry* **17,** 2124–2133.
38. Rudikoff, S. & Potter, M. (1978) *Biochemistry* **17,** 2703–2707.
39. Cook, W. D., Rudikoff, S., Giusti, A. & Scharff, M. D. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 1240–1244.
40. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1979) *Sequences of Immunoglobulin Chains* (U.S. Dept. of Health, Education, and Welfare), Publication No. 80, p. 2008.
41. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 617–619.
42. Barstad, P., Rudikoff, S., Potter, M., Cohn, M., Konigsberg, W. & Hood, L. (1974) *Science* **183,** 962–964.
43. Rudikoff, S., Satow, Y., Padlan, E., Davies, D. & Potter, M. (1981) *Immunochemistry* **18,** 705–711.
44. Kwan, S. P., Rudikoff, S., Seidman, J. G., Leder, P. & Scharff, M. D. (1981) *J. Exp. Med.* **153,** 1366–1370.
45. Rudikoff, S., Potter, M., Segal, D. M., Padlan, E. A. & Davies, D. R. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 3689–3692.
46. Padlan, E. A., Davies, D. R., Rudikoff, S. & Potter, M. (1976) *Immunochemistry* **13,** 945–949.